ADVANCING EARTH AND SPACE SCIENCES

Earth's Future

RESEARCH ARTICLE
10.1029/2024EF004751

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

Uncovering the Dynamic Drivers of Floods Through Interpretable Deep Learning

Yuanhao Xu[1,2,3], Kairong Lin[1,2,3], Caihong Hu[4], Xiaohong Chen[1,2,3], Jingwen Zhang[1,2,3], Mingzhong Xiao[1,2,3], and Chong-Yu Xu[5]

[1]State Key Laboratory for Tunnel Engineering, School of Civil Engineering, Sun Yat-sen University, Guangzhou, China, [2]Guangdong Engineering Technology Research Center of Water Security Regulation and Control for Southern China, Guangzhou, China, [3]Guangdong Key Laboratory of Marine Civil Engineering, Guangzhou, China, [4]School of Water Conservancy Science and Engineering, Zhengzhou University, Zhengzhou, China, [5]Department of Geosciences, University of Oslo, Oslo, Norway

Key Points:

- Proposed a framework for interpretable time series deep learning based on optimal performance to reveal the dynamic drivers of floods
- Unveiled the changes in the dominant driving factors of floods and the roles of driving factors at different moments before a flood
- Elucidated the dynamic responses of floods to the interactions among driving factors

Correspondence to:

K. Lin,
linkr@mail.sysu.edu.cn

Citation:

Xu, Y., Lin, K., Hu, C., Chen, X., Zhang, J., Xiao, M., & Xu, C.-Y. (2024). Uncovering the dynamic drivers of floods through interpretable deep learning. *Earth's Future*, *12*, e2024EF004751. https://doi.org/10.1029/2024EF004751

Received 10 APR 2024
Accepted 10 SEP 2024

Abstract

The formation of floods, as a complex physical process, exhibits dynamic changes in its driving factors over time and space under climate change. Due to the black-box nature of deep learning, its use alone does not enhance understanding of hydrological processes. The challenge lies in employing deep learning to uncover new knowledge on flood formation mechanism. This study proposes an interpretable framework for deep learning flood modeling that employs interpretability techniques to elucidate the inner workings of a peak-sensitive Informer, revealing the dynamic response of floods to driving factors in 482 watersheds across the United States. Accurate simulation is a prerequisite for interpretability techniques to provide reliable information. The study reveals that comparing the Informer with Transformer and LSTM, the former showed superior performance in peak flood simulation (Nash-Sutcliffe Efficiency over 0.6 in 70% of watersheds). By interpreting Informer's decision-making process, three primary flood-inducing patterns were identified: Precipitation, excess soil water, and snowmelt. The controlling effect of dominant factors is regional, and their impact on floods in time steps shows significant differences, challenging the traditional understanding that variables closer to the timing of flood event occurrence have a greater impact. Over 40% of watersheds exhibited shifts in dominant driving factors between 1981 and 2020, with precipitation-dominated watersheds undergoing more significant changes, corroborating climate change responses. Additionally, the study unveils the interplay and dynamic shifts among variables. These findings suggest that interpretable deep learning, through reverse deduction, transforms data-driven models from merely fitting nonlinear relationships to effective tools for enhancing understanding of hydrological characteristics.

Plain Language Summary

The formation of floods is often dynamically influenced by multiple driving factors. Traditional methods based on statistics or hydrological models struggle to clearly understand flood mechanisms due to their limitations. Although deep learning has become an effective tool for flood modeling, its black-box nature makes it difficult to enrich understanding of flood processes through it. Our proposed interpretable deep learning offers a perspective to unveil the dynamic drivers of floods by extracting patterns from deep learning in a reverse deduction approach. We started with models that perform best on flood extremes and identified the dominant factors and variations of floods across 482 watersheds in the United States. These variations exhibit regional characteristics, with precipitation playing a more significant role and showing more pronounced trends in watersheds where floods are primarily driven by precipitation. We found that variables from 2 days before a flood could have a greater impact than those from the previous days. Furthermore, we revealed the combined impact of variables on flooding through deep learning, showcasing their dynamic changes. Interpretable deep learning explores a new avenue to derive new hydrological insights from inverse data, helping hydrologists better understand natural physical processes.

1. Introduction

Flood disasters have always been one of the most significant natural threats to human survival and social development (Tellman et al., 2021; Winsemius et al., 2016). With the intensification of global warming trends, observational evidence and climate model predictions consistently indicate that due to the increased atmospheric water-holding capacity caused by global warming, the frequency of drought and flood disasters is expected to significantly increase (Han et al., 2024; Rentschler et al., 2023; Singh & Basu, 2022). Furthermore, the probability of occurrence of major disasters under extreme scenarios and the uncertainty of their spatial and temporal distribution are gradually increasing (Donat et al., 2016; Hirabayashi et al., 2013; Yin et al., 2021). Extensive research shows that the frequency, seasonality, and scale of global floods vary significantly among different regions, highlighting the urgent need to improve and enhance flood forecasting to gain a more comprehensive understanding of the processes and driving factors behind global watershed flooding (Alfieri et al., 2017; Berghuijs et al., 2016; Bloeschl et al., 2019; Wang et al., 2024).

The formation of floods, as a complex physical process, involves the interaction between hydrological, meteorological and physical characteristics of the watersheds, and can be triggered by multiple mechanisms (Berghuijs et al., 2019; Jiang et al., 2024; Mallakpour & Villarini, 2015; Wasko & Nathan, 2019). For instance, floods can be induced by increased heavy precipitation brought about by climate change; rising temperatures can affect the dynamics of snow accumulation in cold regions, leading to more extreme rainfall events, which might make watersheds dominated by snowmelt more susceptible to extreme precipitation, thereby altering the seasonality and scale of regional floods; long-term changes in land cover can alter soil moisture, evapotranspiration, and land-atmosphere flux exchanges, thus affecting the flood generation process (Anderson et al., 2022; Bertola et al., 2021; Rentschler et al., 2023; Sharma et al., 2018; Sikorska et al., 2015; Westra et al., 2013). Past studies that have inferred the mechanisms of flood generation in watersheds often identify a single dominant mechanism driving the floods, which can only explain partial characteristics of local floods (Do et al., 2020; Kiptum et al., 2024; Merz et al., 2012; Stein et al., 2021). However, a single flood event within a watershed can also be produced through different mechanisms, as the processes affecting the magnitude, frequency, and timing of floods can occur in parallel and interact in various ways across spatial and temporal scales (Hall et al., 2014; Musselman et al., 2018; Slater et al., 2021). The dominant flood mechanisms in a watershed under climate change are not always constant and may change over time. The drivers of floods are dynamically variable in both time and space (Dethier et al., 2020; Regonda et al., 2005; Vormoor et al., 2016).

Current research on the driving factors of floods primarily focuses on classifying flood processes based on events (Berghuijs et al., 2019; Stein et al., 2020; S. Zhang et al., 2022; C. Zhang et al., 2022). Traditionally, flood types have been categorized into four main groups: those induced by intense rainfall leading to maximum flood inundation, those caused by excessive soil moisture resulting in maximum flood volumes, those triggered by snowmelt leading to maximum flood volumes, and those induced by rain-on-snow events causing maximum flood inundation (Berghuijs et al., 2016; Keller et al., 2018; Turkington et al., 2016). These event-based flood types are used to identify the predominant flood category for each watershed (Mallakpour & Villarini, 2015; Slater & Villarini, 2016; Tarasova et al., 2020). Interpretations of historical flood trends and predictions of future floods are generally based on statistical methods using rainfall-runoff data, or by employing process-based models that describe precipitation partitioning at the watershed scale (Archfield et al., 2016; Zhang et al., 2016). Statistical methods, which are largely based on simpler linear regression theories, struggle to extract complex nonlinear relationships related to floods from a vast array of hydrological variables (Bloeschl et al., 2019; Liu, Gui, et al., 2022; Tarasova et al., 2023). Process-based modeling approaches, which rely on complex mathematical formulas and precise understanding of hydrological processes, reflect hydrological processes with many uncertainties and generalized phenomena (Bertola et al., 2021; Do et al., 2017). Attribution faces various challenges that depend on the variables considered. In extreme hydrology, disentangling multiple driving factors can be highly complex (Rong et al., 2024; Slater et al., 2021). A major unresolved issue in hydrology is the lack of consideration for non-stationary drivers on appropriate temporal and spatial scales (Bertola et al., 2021; Machado et al., 2015). If the attribution framework is too narrow, failing to consider multiple plausible driving factors, significant drivers of hydrological changes may be overlooked (Guo et al., 2017; Whitfield et al., 2012). Due to the complexity and dynamic nature of flood generation mechanisms, the effectiveness and reliability of both approaches are limited by their ability to represent the relevant processes controlling flood responses (Merz et al., 2021; Slater et al., 2021).

In recent years, deep learning has gradually become an effective approach for simulating and discovering hydrological patterns due to its powerful capability to extract complex nonlinear relationships between variables (Kratzert et al., 2018; Nearing et al., 2024; Xu et al., 2023). Deep neural networks such as LSTM, TCN, and GRU have achieved excellent performance in areas like flood processes, water quality, and soil moisture variations, as confirmed by a large amount of recent research (J. Zhang et al., 2018; Kratzert et al., 2019; Feng et al., 2020; Frame et al., 2022). However, the "black box" nature of these data-driven models results in a lack of interpretability, making it unclear how features influence outputs and their directions, with poor visibility into feature importance (Lees et al., 2022; Shen, 2018). Moreover, solely using them does not enhance our understanding of hydrological processes (Shen et al., 2023; Tripathy & Mishra, 2024). Understanding the underlying principles behind the decisions of deep learning models is crucial not only for bolstering our confidence in their use (i.e., building trust) and for further improving the models (i.e., troubleshooting) but also essential for discoveries in hydrological science (Cheng et al., 2022; Feng et al., 2023; Jiang, Zheng, et al., 2022; Y. Liu, Bian, & Shen, 2023). This leads to the question of how to uncover regional spatiotemporal hydrological features from machine learning models trained on flood processes, which could provide new insights into hydrology.

Recent studies have advanced the field by exploring the correlation between input rainfall-runoff features and the decisions of deep learning models, a practice known as interpretable deep learning (Lusch et al., 2018; Mamalakis et al., 2023; Murdoch et al., 2019; Shen, 2018). Most existing research using interpretable techniques has relied on simple tree regression models to reveal the overall climatic contributions of watersheds to flood generation (Konapala et al., 2020; Lundberg et al., 2020; Stein et al., 2021). Even the few explanations applied to deep learning are limited to calculating feature importance, struggling to reveal the dynamic changes in driving factors for individual flood events (such as interannual variations) and the interactions between factors. It's noteworthy that using deep learning to simulate the time series of flood formation and achieving satisfactory simulation results is a prerequisite for applying interpretable methods (Jiang, Bevacqua, & Zscheischler, 2022; Jing et al., 2023; Q. Liu, Gui, et al., 2022). Poor simulations inevitably lead to significant errors, especially in predicting peak floods, and explanations based on such simulations would deviate from reality (Bian et al., 2023; Koya & Roy, 2023; Liu, Bian, & Shen, 2023). Previous explorations using LSTM have encountered gradient problems in recurrent neural networks, which tend to give recent variables significant impact (S. Anderson & Radic, 2022; De la Fuente et al., 2024). Moreover, in most cases, LSTMs struggles to accurately predict flood peaks, introducing potential bias into the analysis of flood driving factors. However, the latest studies based on Transformer variants like Informer, which utilize self-attention mechanisms, are more sensitive to peaks and offer a possibility for accurately identifying driving factors (Zhou et al., 2021).

In this study, our objective is to construct a framework for interpretable deep learning to reveal the dynamic driving factors of floods. We have trained the Informer model on 482 watersheds in the United States, and comparisons with other deep learning models ensure that the Informer possesses the most reliable flood peak forecasting capability. SHAP interpretability techniques were employed to quantify the flood response to hydrometeorological variables, identify the dominant driving factors of watershed floods, and elucidate the mechanisms of these variables at different moments preceding flood events. We also examined the abrupt changes and trends of dominant driving factors over time and revealed the dynamic changes under different flood generation mechanisms. Furthermore, we delved into and quantified the flood response to interactions between variables and associated the abrupt changes in dominant driving factors to reveal their dynamic changes.

2. Methods

2.1. Informer

Informer is a deep learning framework based on Encoder-Decoder and self-attention mechanism. It is designed to process large-scale, complex and irregularly sampled time series data, and can capture complex long-term dependencies in the data, which is very important for large-sample hydrological data (Zhou et al., 2021). It is the key to feature extraction. It introduces two special structures, ProbSpare self-attention and self-attention distillation, to improve the shortcomings of slow calculation and limited memory of the attention layer in the traditional Transformer structure (L. Shen & Wang, 2022). The former ProbSpare self-attention will amplify the effective attention score to reduce spatiotemporal complexity, and the latter self-attention distillation reduces memory consumption by shortening the length of each layer input (Zhou et al., 2023). In addition, the parallel generative decoder mechanism in Informer implements a forward calculation to output all prediction results for long-term sequences instead of making predictions in a step-by-step manner, which greatly improves the inference speed of long-sequence predictions (Kroner et al., 2020; C. Zhang, Zhou, et al., 2022). The structure of Informer is shown in Figure 1.

The input data is first processed through the operation of fixed positional embedding to preserve the local context. After the encoding step, the data format that enters the Encoder layer is as follows:

10.1029/2024EF004751

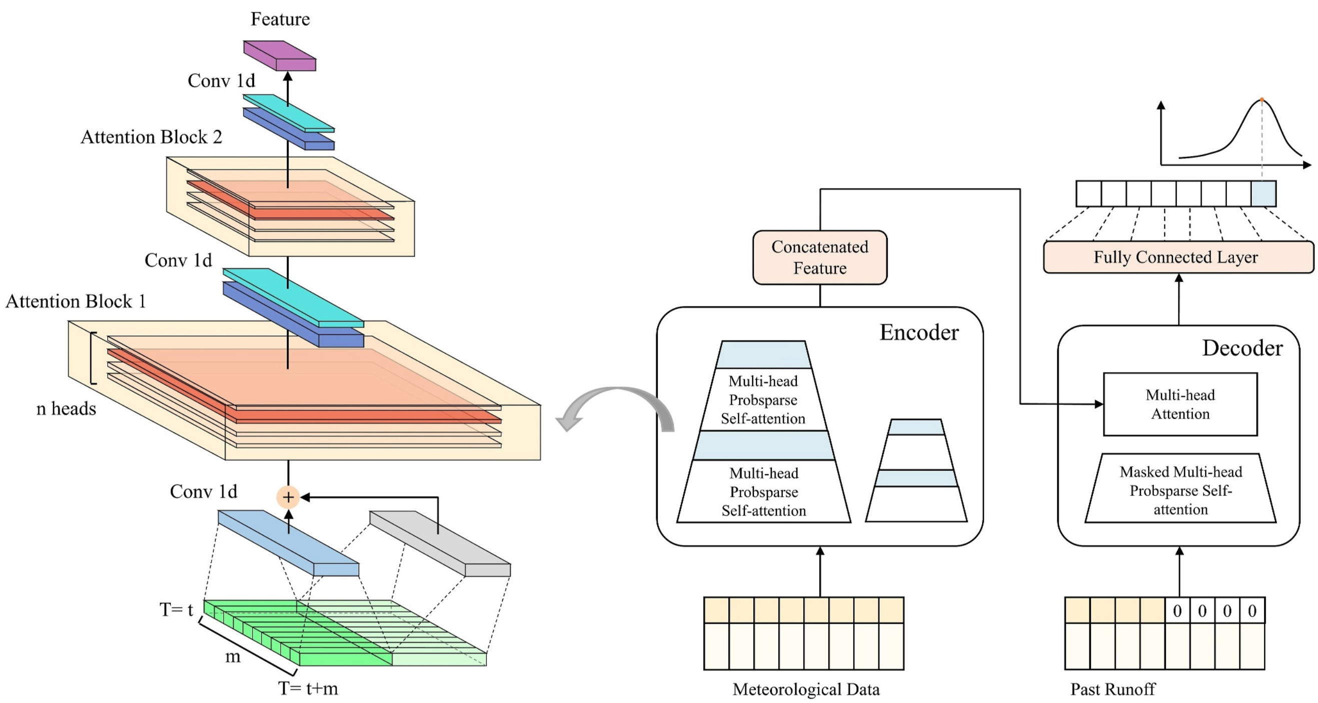

Figure 1. Encoder and Decoder Structure of the Informer for Simulating the Flood Process. The ProbSparse self-attention in the Encoder employs KL divergence to measure the similarity between probability distributions and q, distinguishing important queries. It also generates diverse sparse query-key pairs, thereby avoiding severe information loss.

$$X^t_{\mathrm{feed}[i]} = \alpha u^t_i + PE_{(L_x\times(t-1)+i)} + + \sum_p \left[SE_{(L_x\times(t-1)+i)}\right]_p \tag{1}$$

where u represents the vector of input data, $i \in \{1, \ldots, L_x\}$, with L_x being the length of the sequence. α is a factor that balances the magnitude between scalar projection and local/global embeddings, and is set to 1 when the data has been normalized.

ProbSparse self-attention is introduced, utilizing the KL divergence to measure the similarity between probability distributions and q to differentiate important queries. The sparsity of the ith query can be represented as follows:

$$M\left(q_i, K\right) = \ln \sum_{j=1}^{L_K} e^{\frac{q_i k_j^T}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{q_i k_j^T}{\sqrt{d}} \tag{2}$$

where d is the dimension of the mapped input sequence, and L_K is the length of the sequence. Each key is only allowed to attend to the dominant queries, thus the ProbSparse self-attention can be represented as:

$$A(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{3}$$

Subsequently, by utilizing distilling, dominant features with substantial attention are assigned greater weights, and dilated convolutions are added between each encoder and decoder layer, significantly reducing the temporal dimension of the input. This operation can be represented as:

$$X^t_{j+1} = \mathrm{MaxPool}\left(ELU\left(\mathrm{Conv1}d\left(\left[X^t_j\right]_{AB}\right)\right)\right) \tag{4}$$

10.1029/2024EF004751

Where X_{j+1}^{t} is the output of the multi-head ProbSparse self-attention layer, $[\ ]_{AB}$ is the result of the previous multi-head ProbSparse self-attention layer, and $ELU()$ is the activation function of Conv1d() along the temporal dimension.

In the decoder section, the Informer employs a batch generation approach to output multi-step predictions in one go to enhance the running speed, which can be represented by the following computation:

$$X_{de}^{t} = \mathrm{Concat}\left(X_{\mathrm{token}}^{t}, X_{0}^{t}\right) \in R^{(L_{\mathrm{token}}+L_{y})\times d_{\mathrm{model}}} \tag{5}$$

Where X_{token}^{t} is the start token, which is dynamically sampled from a portion of the input sequence close to the prediction target, and X_{0}^{t} is the placeholder for the predicted sequence. L_{token} and L_{y} are the lengths of the start token and the predicted sequence, respectively, with d_{model} being the dimension.

2.2. SHAP

SHAP (SHapley Additive exPlanation) is a machine learning interpretability method inspired by cooperative game theory (CGP), designed to reveal the impact of each input feature on individual predictions (Štrumbelj & Kononenko, 2014). As the computational capacity and complexity of machine learning models continue to advance, understanding the internal workings of these models and how they make decisions becomes increasingly challenging (Fong & Vedaldi, 2017; Rudin, 2019). Enhancing the interpretability of "black box" models allows for better understanding of their predictions, and increases the generalizability and credibility of machine learning algorithms. The core idea of the SHAP method originates from the Shapley value, which is used to allocate the contribution of each participant to the total gain in cooperative games—a concept borrowed from game theory. In SHAP analysis, each prediction in deep learning is viewed as a game to determine the marginal contribution of each feature across different combinations (Aas et al., 2021; Stojic et al., 2019). A positive SHAP value indicates a positive impact of the feature on the prediction, while a negative SHAP value indicates a negative impact.

Compared to traditional feature importance analysis, SHAP holds an advantage by not only presenting the global importance of features but also explicitly detailing the specific impact of each feature on individual predictions and the complex interactions between features (Li et al., 2022). This aids in understanding the predictive process and decision rationale of complex "black-box" models. For deep learning time series predictions, SHAP values can be derived for both variable dimensions and temporal dimensions, providing the possibility of revealing more precise flood driving factors. The SHAP values is calculated as:

$$\phi_{i}(f, x) = \sum\nolimits_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}\left[f(x_{S}\cup\{x_{i}\}) - f(x_{S})\right] \tag{6}$$

where $\phi_{i}(f,x)$ represents the SHAP value of feature x_{i}, f denotes the predictive function of the model, N and S are the sets encompassing all features and the set excluding x_{i} respectively. x_{S} signifies the input under the given feature set S, and $|N|$ and $|S|$ correspond to the sample count of sets N and S respectively.

For this work, we describe an approach to perform interpretable deep learning rainfall-runoff time series simulations on a large global sample to reveal the dynamic drivers of flooding. This process can be represented by Figure 2. A hydrometeorological time series data set from a large sample of Caravan catchments was used to train the Informer model to build non-linear prediction maps from meteorological factors (i.e., precipitation, temperature and day length) and hydrological factors (soil moisture) to daily flow. Accurate flood process simulation is the prerequisite and key to ascertaining the potential hydrological laws contained in deep learning. It ensures that the working principles derived from deep learning are objective and reliable. The advantage of using Informer lies in its strong sensitivity to extreme values and good capture properties of long-distance dependencies, which is beneficial for flood estimation and can generate accurate contributions of input variables to flood peaks at different times in the past. Physically realistic mapping from inputs to outputs helps derive hydrologically meaningful insights from subsequent model interpretation. After that, SHAP interpretation technology is used to interpret the trained Informer to reveal the dynamic response of floods to variables. Informer's training mechanism is based on inferring target values from several past moments, which are defined as time steps, forming a three-dimensional array with the number of variables and batch size for deep neural network learning. The occurrence of sustained rainfall and prolonged high soil moisture prior to a flood significantly impacts flood

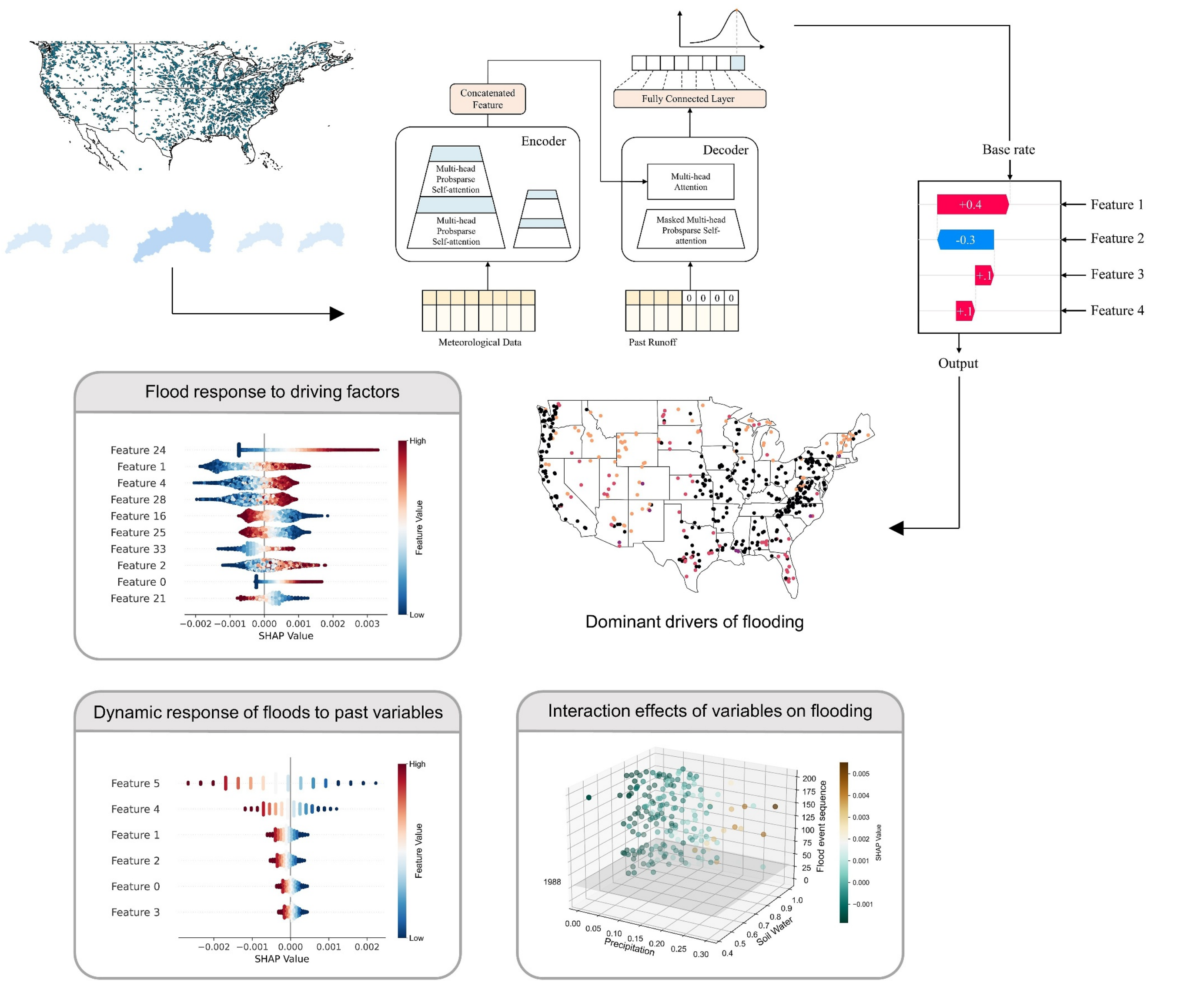

Figure 2. Interpretable Deep Learning Framework Revealing the Dynamic Drivers of Floods. Exploring the inner workings of well-trained Informer models to investigate the dominant driving factors of floods, the flood's response to variables at different past times, and the dynamic interactions among variables.

generation. Information regarding these factors is encapsulated within the three-dimensional arrays utilized for each prediction. Interpretation over time steps supports quantifying the contribution of variables at different times to flood generation. Flood drivers behave differently across events over time and may potentially change. We count the occurrence times of all dominant driving factors of flood events based on a 10-year sliding window, and the MK test is used to verify whether they have mutated every 10 years, that is, the mechanism of flood generation has undergone fundamental changes. Finally, we calculate the interaction SHAP value for the Informer trained in the first step, reveal the interaction between dominant driving factors and other variables on floods at time steps, and explore the reasons for changes in flood generation mechanisms.

3. Data

This study considers the Caravan data set, which includes daily meteorological forcing data and daily river flow observations for 482 watersheds in the United States, from the latest large-sample research. These catchment areas are minimally influenced by human activities and are suitable for the time series simulation of deep learning (Addor et al., 2017; Kratzert et al., 2024). Our analysis is confined to catchment areas with at least 20 years of

10.1029/2024EF004751

Table 1
Selected Daily Meteorological and Hydrological Variables

Group	Feature name	Feature description	Aggregation	Unit
Precipitation	total_precipitation	Precipitation	Daily sum	mm/day
Evaporation	potential_evaporation	Potential evaporation	Daily sum	mm/day
Temperature	temperature_2m	Air temperature	Daily min/max and mean	°C
	dewpoint_temperature_2m	Dew point temperature	Daily min/max and mean	°C
Snow	snow_depth	Snow water equivalent	Daily min/max and mean	mm
Soil Water	soil_water_layer_1	Soil water volume 0–7 cm	Daily min/max and mean	m^3/m^3
	soil_water_layer_2	Soil water volume 7–28 cm	Daily min/max and mean	m^3/m^3
	soil_water_layer_3	Soil water volume 28–100 cm	Daily min/max and mean	m^3/m^3
	soil_water_layer_4	Soil water volume 100–289 cm	Daily min/max and mean	m^3/m^3
Radiation	surface_net_solar_radiation	Shortwave radiation	Daily min/max and mean	Wm^{-2}
	surface_net_thermal_radiation	Net thermal radiation at the surface	Daily min/max and mean	Wm^{-2}
Wind	u_component_of_wind_10m	Eastward wind component	Daily min/max and mean	ms^{-1}
	v_component_of_wind_10m	Northward wind component	Daily min/max and mean	ms^{-1}

discharge records between 1981 and 2020, to ensure a sufficient sample size for training deep learning models. The temporal series attributes can be represented by Table 1. The data for each watershed is retrieved from the Caravan data set, which provides daily hydrometeorological time series, mostly spanning from 1981 to 2020. For these catchment areas, the sample size of daily discharge records ranges from 6,820 to 14,236, with a median time step of 10,287. Furthermore, with future updates to Caravan in more watersheds, the method we propose could potentially be validated more extensively, which is also one of the reasons for our choice of Caravan.

The programming language of choice is Python 3.9, the libraries used for preprocessing and managing our data are Pandas, NumPy, and PySwarms. We use the Pytorch deep-learning framework and the NVIDIA RTX 3090Ti GPU to train the models.

We employed the Peak Over Threshold (POT) method to extract flood peaks and corresponding predictor variables from the flow time series of each watershed (Lang et al., 1999). To ensure that each flood peak is relatively independent, the minimum time interval between two flood peaks was set to 5 days. Additionally, another constraint was to ensure that the intermediate flow between two consecutive flood peaks is less than 75% of the lower of the two flood peaks. The number of flood peaks extracted in this manner is determined by the threshold in the POT method (M. Shen & Chui, 2023).

To enable the predictor variables to be recognized and used for training by the Informer model, we converted these data into a format understandable by the model, that is, divided into input samples and corresponding labels. This conversion process involves organizing the data into a sliding window format with time series features, facilitating model training and subsequent forecasting tasks. This sliding window method constructs multidimensional array-form observations using consecutive time point data to predict the flow value at a specific time point t, with the model referring to the feature values of the previous n time points (i.e., t-1, t-2, …, t-n), where n represents the time steps (Kratzert et al., 2021). Subsequently, the data is normalized to ensure that the numerical range is between 0 and 1. This step is particularly crucial for gradient-based deep learning algorithms, as these algorithms are highly sensitive to the scale of input data and often require normalization to ensure algorithmic convergence. Finally, a denormalization step was implemented to rescale the model's output data to obtain accurate flow prediction values.

The hyperparameters such as the number of encoder and decoder layers, the number of attention heads, the factor in ProbSparse self-attention, and the sliding window size are configurations of the Informer model that determine the efficiency and accuracy of flood estimation (Zhou et al., 2021). The number of encoder and decoder layers dictates the depth of the model, with more layers enhancing the model's capability to process complex data. Multiple attention heads allow the model to learn information in parallel across different representational subspaces, handling the hydro-meteorological features of the input with multiple sets of self-attention. The factor in

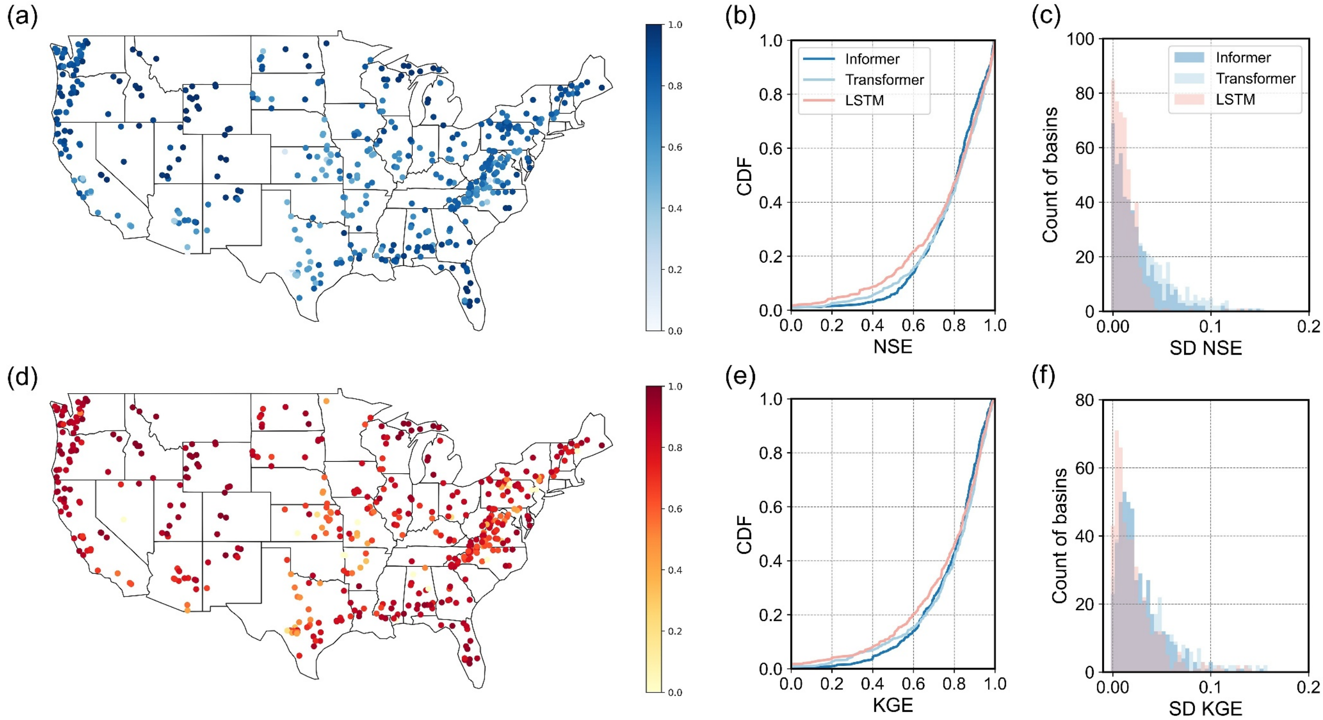

Figure 3. Performance of rainfall-runoff simulation with Informer using 10-fold cross-validation and comparison with Transformer and LSTM. (a, d) Average distribution of Nash-Sutcliffe Efficiency (NSE) and KGE for simulation results by Informer; (b, e) average cumulative frequency of NSE and KGE; (c, f) distribution of standard deviation values for NSE and KGE across 10-fold cross-validation. The NSE and KGE values were calculated using all samples in respective testing data sets.

ProbSparse self-attention can reduce computational complexity by making the attention mechanism sparse and focusing on key parts of the sequence. The sliding window size represents the timestep length of input data that the Transformer uses for each flow prediction. Simulations in each watershed require as suitable hyperparameters as possible (Zhou et al., 2023). For a fairer comparison, we use the Tree-structured Parzen Estimator (TPE) method for automatic hyperparameter optimization (Ozaki et al., 2020). It builds a surrogate model using past observations to predict the value of the objective function at any point in the hyperparameter space. During this process, a Parzen window estimator (also known as kernel density estimation) is used to construct the surrogate model. The TPE algorithm divides the hyperparameter space into two parts, one corresponding to better objective function values and the other to worse values. It infers the next most likely hyperparameter point to improve performance by comparing the probability density functions of these two parts.

4. Results and Discussion

4.1. Informer Forecast Performance and Comparison With Benchmarks

Accurate and stable predictions are prerequisites for deriving meaningful hydrological information from deep learning. Using the Informer for daily rainfall-runoff simulations across 482 watersheds in Caravan and based on 10-fold cross-validation calculated on the test set, we assessed the overall performance of the deep learning model in these watersheds and its estimation of peak flows. Figure 3 shows the Nash-Sutcliffe Efficiency (NSE) evaluation metric between observed and simulated streamflow in the test set. According to the results, the Informer was able to effectively reproduce observed stream processes. Although the standard deviation of LSTM is more stable, the Informer achieved better simulations in most watersheds, and the Transformer exhibited greater fluctuations in cross-validation. This indicates that the Informer network architecture faithfully captured the underlying dynamic relationships between runoff-related variables in most watersheds. Across all watersheds, 290 (over 60%) achieved an NSE greater than 0.8. The performance in the majority of watersheds along the western coast and the east significantly exceeded that in the central regions, consistent with other studies using deep learning for rainfall-runoff modeling, which found modeling in arid regions to be more challenging (Addor et al., 2018; Knoben et al., 2020; M. Shen & Chui, 2023).

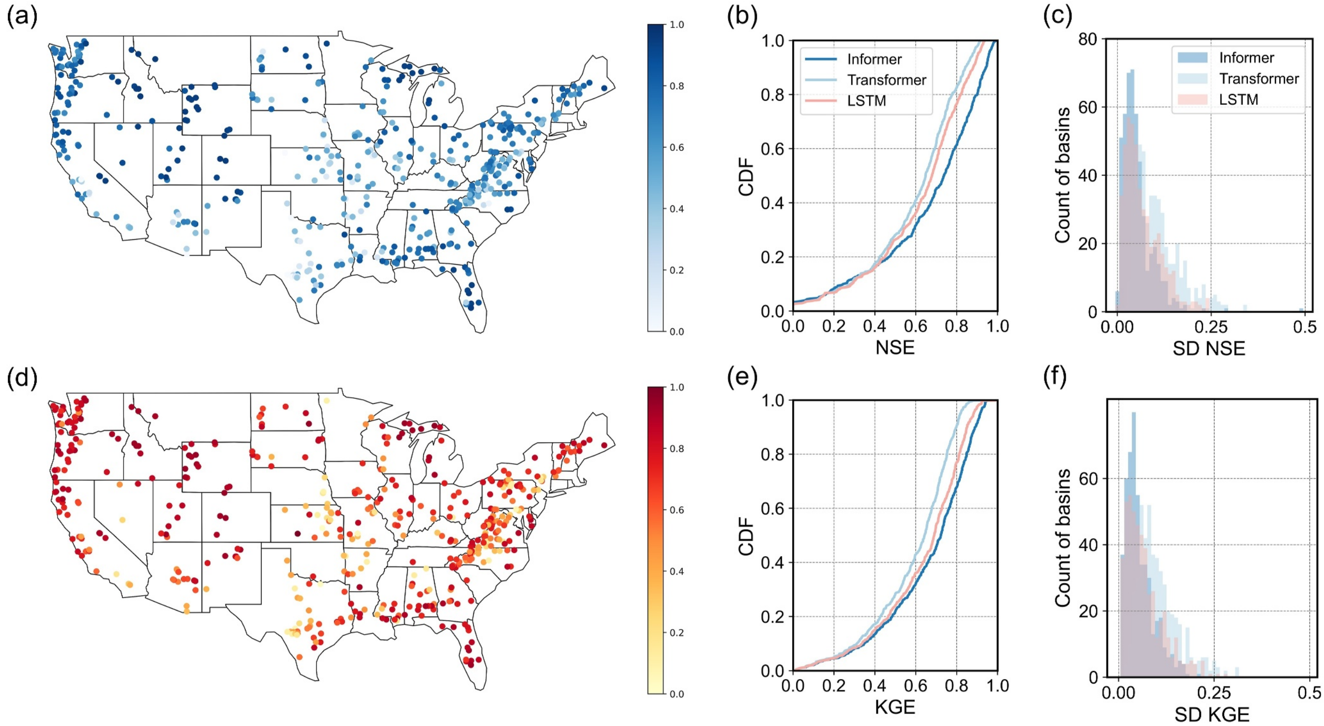

Figure 4. Performance of flood peak simulation with Informer using 10-fold cross-validation and comparison with Transformer and LSTM. (a, d) Average distribution of Nash-Sutcliffe Efficiency (NSE) and KGE for simulation results by Informer; (b, e) average cumulative frequency of NSE and KGE; (c, f) distribution of standard deviation values for NSE and KGE across 10-fold cross-validation.

Using the POT method, a total of 21,605 significant flood peaks were identified at the outlet stations of 482 watersheds, accounting for 0.9% of the total sample (0.4%–1.8% for a single watershed). The performance of the model in reproducing the observed flood peaks is also evaluated and showed in Figure 4. For about 70% of the watersheds, the NSE of flood peaks simulated by Informer exceeds 0.6. However, in some watersheds such as those around the Rocky Mountains in the Middle East of the United States, flood peaks are difficult to accurately estimate (NSE is less than 0.5). However, these watersheds account for less than 18% of all watersheds considered. Compared to other benchmarks, the Informer achieved most robust and reliable control of peak flows in cross-validation. Overall, given Informer's performance in reproducing the streams and flood peaks observed by Caravan, it is considered well-trained and can be used to explore underlying hydrological patterns in large samples. When we subsequently use the interpretable method, we only focus on watersheds with flood peak simulated NSE higher than 0.5, which helps ensure that the obtained flood driving factors and mutation information are meaningful.

4.2. Uncovering the Dominant Dynamic Driving Factors of Floods

In the previous results, we obtained an Informer model trained on a large sample. We calculated the SHAP value at the peak of each flood. The calculation of each flood peak considered the disturbance of flow rate by 36 variables (a total of 216 variable values) at the time step of the past 6 days. The average SHAP value of the variables across all watersheds reflects the dominant drivers of flooding, as shown in Figure 5. Most river floods are controlled by soil water content and are mainly distributed in the eastern and southern coastal areas and the central and western inland areas. This means that excess soil water caused by precipitation and other factors has become the main cause of floods. Flooding in the western coast and eastern inland areas is driven mainly by precipitation, and this spatial distribution pattern is consistent with previous studies (Jiang, Zheng, et al., 2022). In the Rocky Mountains and near the Great Lakes, most flooding comes from snowmelt and temperature, which is consistent with objective facts. Studies have shown that rising temperatures in late spring in these areas cause melting snow to flow into rivers (Bates et al., 2021; Berghuijs et al., 2016).

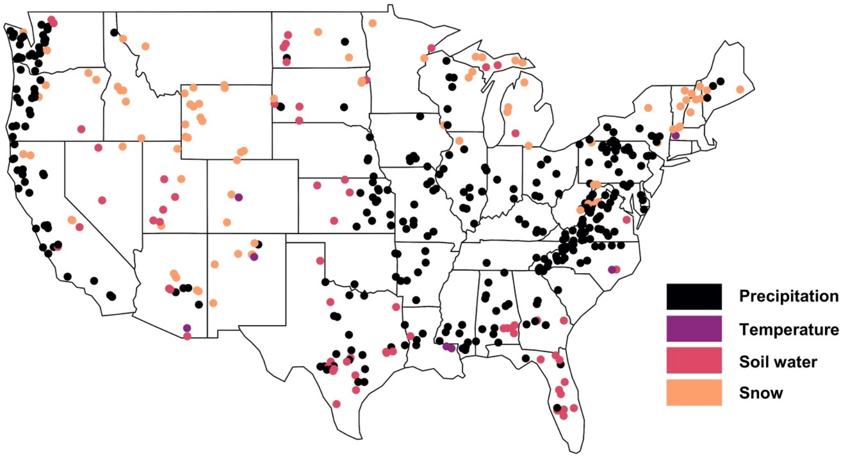

Figure 5. Distribution of dominant flood driving factors revealed by interpretable deep learning. The dominant driving factors of the watersheds are determined by summing and ranking the positive SHAP values for each flood event, where multiple factors within each type are weighted according to their count in the total number of variables.

Variables at different times in the past will have significantly different effects. For example, the snow depth in a period before the flood peak will dominate the generation of runoff, rather than the recent snow accumulation before the flood peak. At the same time, the impact of changes in variable values will also be covered in past time steps. To reveal the response over time steps, we selected three representative watersheds, ensuring that they were spatially evenly distributed and driven by different factors (precipitation, soil water, and snow cover). The graphs in Figure 6 respectively show the effects of each variable in the three watersheds at different values (red and blue represent high and low values respectively), where the order of the effects from high to low is the order of the variables. Furthermore, for the variables that produced the largest flood response in each watershed, we describe the statistics of their SHAP values for each day at the Informer's time step. For watershed (a), soil water acts as the dominant driver of floods, promoting flood peaks at high values and suppressing floods at low values. Super seepage runoff resulted from saturated soil and water infiltration from unsaturated soil play a key role in the generation of floods. The effect of soil water content on the day before the flood is far greater than that of other times in the time step, and decreases from the closer past to the further past. The same result was also seen in watershed (b), which was subject to rainfall-dominated floods, in which the rainfall on the day before the flood became the most critical factor. For the watershed (c), the snow cover 2 days before the flood exhibited the most significant positive impact on the flood behavior, whereas the snow cover one day before the flood had a relatively smaller impact. This indicates that the snow cover in the selected watershed has a delayed effect on flood generation.

To investigate SHAP values and the impact of time steps in watersheds dominated by different factors, we calculated the SHAP value proportions of the dominant factors among all variables (Figures 7a, 7c and 7e). The size of the points represents the number of flood events within each watershed. In most watersheds where floods are dominated by precipitation, precipitation exerts a more significant control on flooding relative to other factors, particularly in the western coastal and eastern inland regions where the SHAP value proportion exceeds 60%. In contrast, the control of precipitation is weaker in the northeastern region, with SHAP value proportions ranging between 40% and 50%. For watersheds where floods are dominated by soil moisture, regional differences are more pronounced, with the most significant control observed in the southern coastal areas, where the SHAP value proportion of soil moisture exceeds 70% in some watersheds. In watersheds where floods are dominated by snow, although the number of flood events is relatively lower in the Rocky Mountain regions compared to the Great Lakes regions, the impact of snow is more significant, possibly due to the perennial snow cover in the former, making snowmelt a major source of increased river water volume. Additionally, we calculated the relative SHAP values and cumulative distribution function curves for the 6 days preceding the flood events (Figures 7b, 7d and 7f). The results indicate that the flood response to past variables varies across watersheds with different dominant factors. Specifically, in precipitation-dominated watersheds, the relative SHAP value for precipitation from the previous day far exceeds that of other times, indicating that precipitation rapidly triggers flooding. In watersheds dominated by soil moisture, the influence extends further back in time, not being solely affected by the

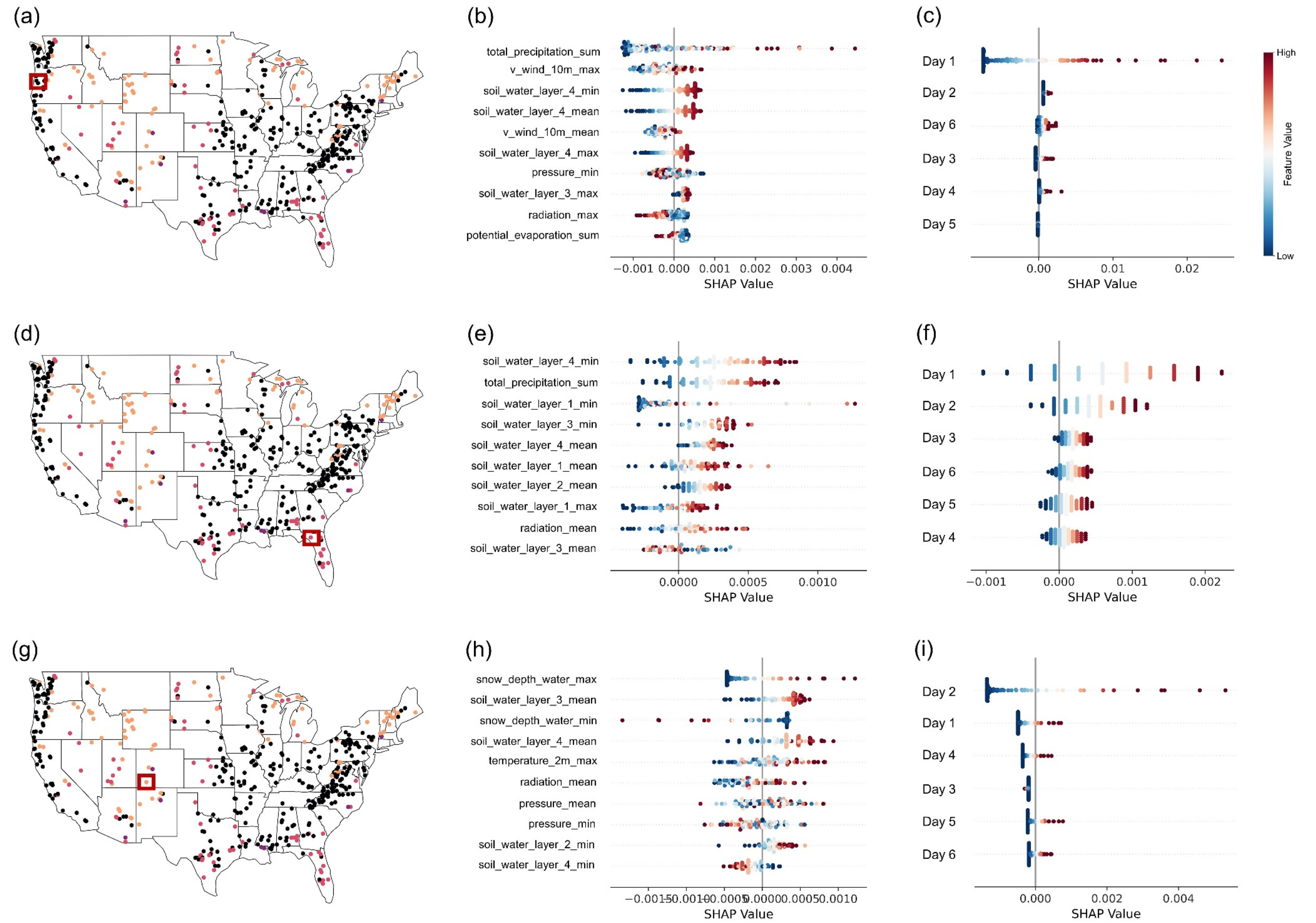

Figure 6. (a, d, and g) Geographic locations of three representative watersheds; (b, e, and h) Role and sequence of driving factors at different values as revealed by interpretable techniques; (c, f, and i) Dynamic response of floods to dominant driving factors at various time steps before the event. Day 1–6 denotes the time from closest to furthest from the occurrence of the flood.

precipitation of the previous day. This phenomenon is even more pronounced in snow-dominated watersheds, where the cumulative distribution function curve is more evenly distributed, suggesting that floods are more influenced by snowmelt from the past 2–6 days.

The dominant driving factors of floods may change over time, and relying on traditional methods or identifying a single driving factor over long time series often fails to reflect the objective reality. We calculated the dominant driving factor for each flood event to assess its temporal changes. A sliding window of 10 years was used to extract all floods within the window. The overall dominant driving factors and their occurrences every 10 years were subjected to the Mann-Kendall test to determine the presence of abrupt changes. The timing and trends of these changes are illustrated in Figure 8a. To explore the extent of reversal in dominant driving factors, we calculated the p-values for different dominant driving factors experiencing abrupt changes across all watersheds (Figure 8b). It reveals that the watersheds experiencing mutations are primarily distributed along the eastern and western coasts, with mutation times in the eastern coastal regions concentrated between 1981 and 1990, and those in the eastern inland regions concentrated between 1990 and 2000. The mutation times in the northern watersheds along the western coast occur later than those in the southern watersheds. The trend of dominant factors exhibits regional characteristics. In the northeastern region, an increasing trend in major factors predominates. This increase is particularly significant in the precipitation-dominated watersheds of the Upper Mississippi River. Conversely, a decreasing trend is predominant in the Midwest. Watersheds dominated by precipitation showed more significant changes than those dominated by soil water, indicating the latter's greater stability over time. We calculated the dominant driving factors for each flood event over a 10-year sliding window for three

10.1029/2024EF004751

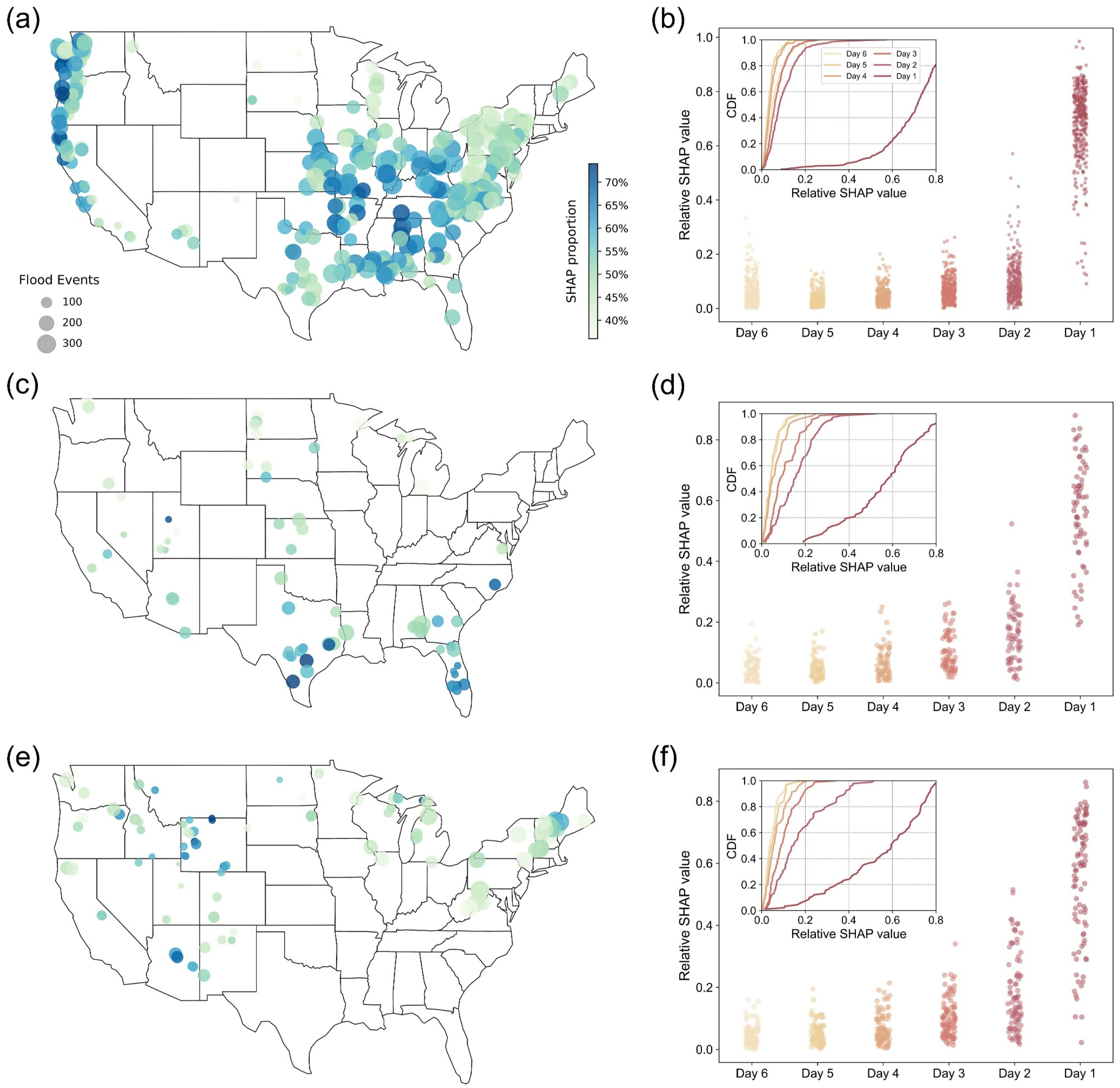

Figure 7. (a, c, and e) SHAP proportion of the dominant factors in the watershed where the dominant factors of floods are rainfall, soil water and snowmelt. The size of the bubble reflects the number of flood events. (b, d, and f) Relative SHAP values and cumulative distribution function curves of the past 6 days in the watershed where the dominant factors of floods are rainfall, soil water and snowmelt.

representative watersheds to observe changes in flood events dominated by different factors (shown in Figures 8c, 8d and 8e). For watershed (a), precipitation-dominated flood events are on an increasing trend, with an overall tendency toward more frequent flooding within the watershed. Watersheds (b) and (c), controlled by soil water and snowpack respectively, show a decreasing trend in the dominance of these factors every 10 years, with a significant decrease in soil water-controlled floods in watershed (b), accompanied by a reduction in overall flood events. This suggests that the control of dominant driving factors changes over time and varies among different types of dominant driving factors.

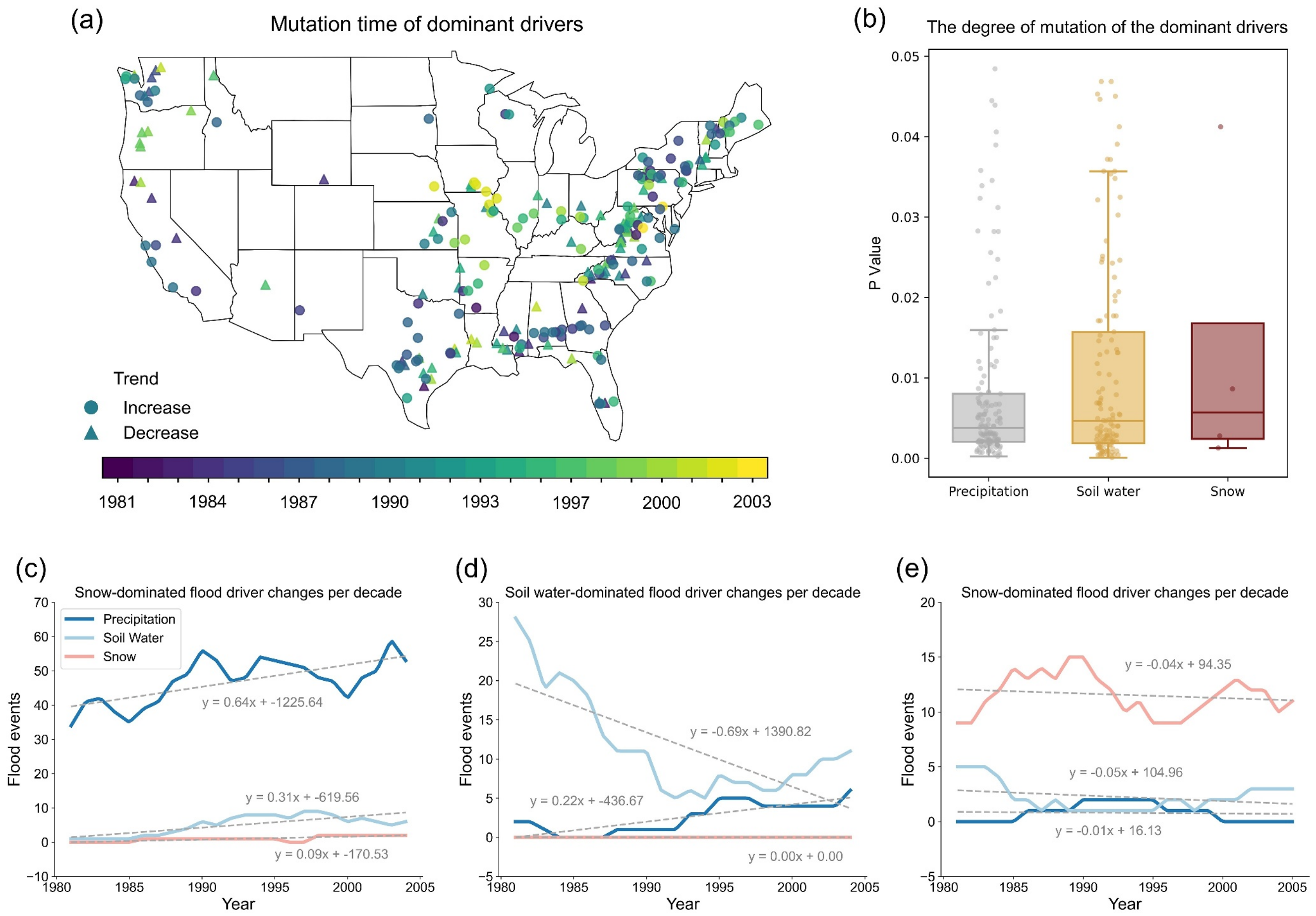

Figure 8. (a) Distribution of watersheds with abrupt changes in dominant flood driving factors and the timing of these changes; (b) Extent of abrupt changes in different types of dominant driving factors, a smaller *p*-value denotes a more significant change; (c–e) Changes in flood driving factors per decade in three representative watersheds. The dashed line represents the trend line fitted based on the number of floods occurring due to driving factors per decade.

4.3. Dynamic Interactions of Flood Drivers

The formation of floods is often influenced by multiple driving factors, or composite factors, which result in differences in the flood formation process across various times and regions (Merz et al., 2022; Slater et al., 2021). The interaction between these composite elements needs to be revealed. For example, floods caused by rapid melting due to strong sunlight versus slow melting of ice caps, and differences in runoff patterns due to deep and shallow soil moisture (S. Zhang, Zhou, et al., 2022). We categorized all considered variables into three types: precipitation, soil moisture, and snowpack, and calculated their interaction SHAP values in each watershed. The overall interaction is represented by the sum of SHAP interaction values (Figures 9a, 7c and 7e), where the color reflects the magnitude of the variable values in the column, vertical dots represent the stacking of flood events, and the horizontal range of dots indicates the strength of the interaction. The variable interactions for each flood event can be depicted in a graph, with variable values and SHAP values under interaction represented by the vertical axis and color, respectively. To reveal the overall extent of interactions between variables across all flood events, we calculated the total interaction SHAP values between pairs of variables (Figures 9b, 7d and 7f). In watershed (a), the points for precipitation and soil moisture (outlined in gray) received a broader range of assignments, with the average SHAP interaction value exceeding 0.2 in the figure, indicating a strong interaction between them in watershed (a). In watersheds (b) and (c), soil moisture and precipitation, snowpack, and temperature respectively showed significant interactions.

The impact of multiple driving factors varies with different values of variables and also shows different effects on flood events over time. We extracted the previously most significant variable group to reveal the dynamic interactions between variables (Figure 10). For watershed (a), which is dominated by precipitation, a promotion

10.1029/2024EF004751

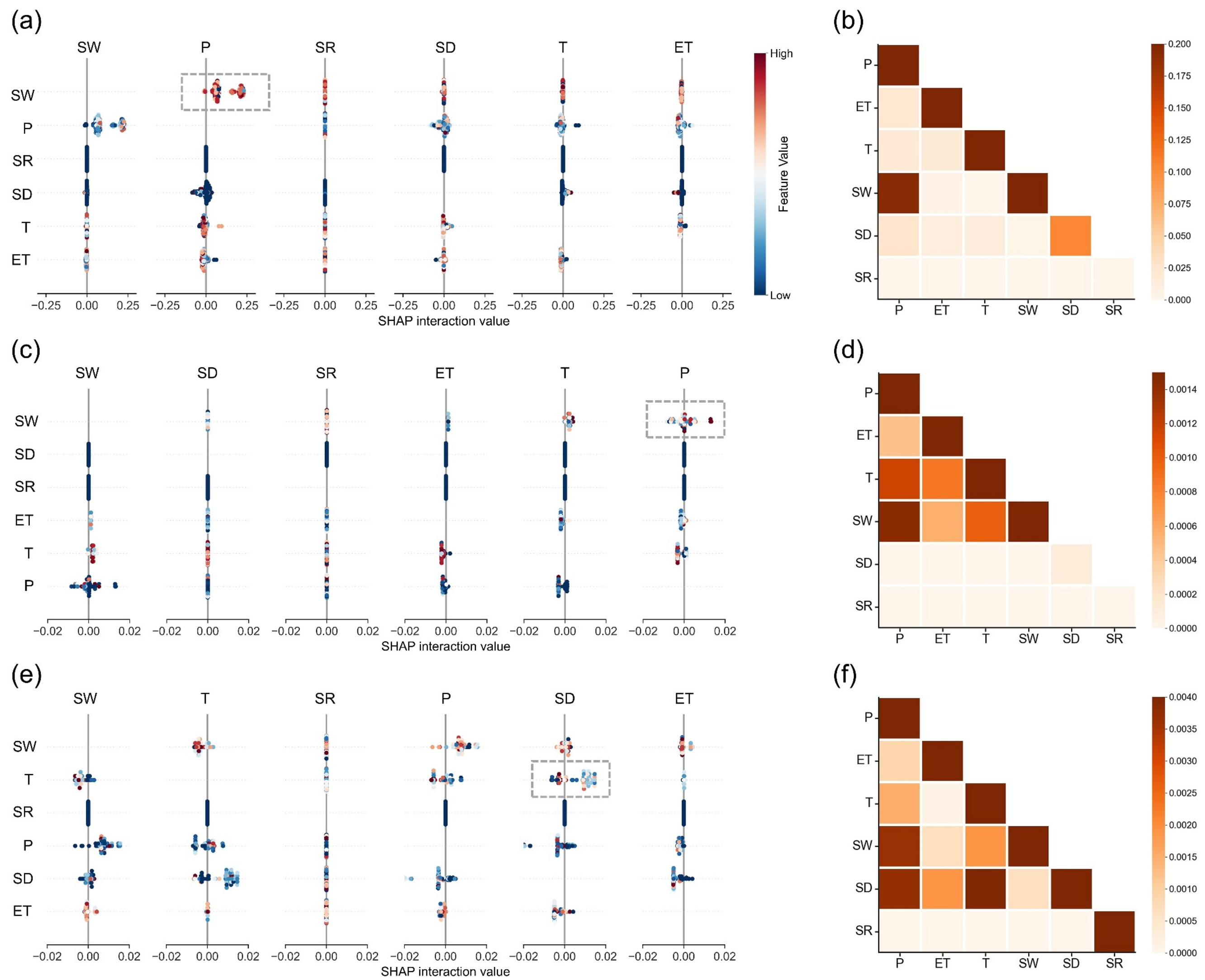

Figure 9. (a, c, and e) Pairwise interactions of different driving factors in each flood event across three watersheds. P, T, ET, SW, SR, and SD represent precipitation, temperature, potential evaporation, soil water, solar radiation, and snow depth, respectively. The most significant interactions are highlighted by gray bounding boxes; (b, d, and f) Overall pairwise variable interactions across three watersheds are colored by interaction SHAP values, with the most significant being P-SW, SW-P, and SD-T respectively.

effect on floods is usually observed only when precipitation exceeds a certain threshold and soil moisture is high. At the same time, the number of interaction effects that promotes floods after the mutation point in 1988 has increased over time. For watershed (b), under the same soil moisture conditions, significant precipitation increases the likelihood of floods, and after 1999, floods controlled by soil moisture and precipitation have noticeably decreased. In watershed (c), flood periods are often accompanied by lower temperatures and greater snow accumulation, and after 1984, floods promoted by interactive effects have sharply decreased, shifting toward more inhibitory interactions instead. These results reveal the dynamic interactions between variables and explore potential factors driving flood change from the data.

In recent decades, the driving factors of watershed floods have been potentially changing under climate change, and the hydrological processes in corresponding flood events are not static (Davenport et al., 2020; Hall et al., 2014). The long-term regularity of flood formation mechanisms is difficult to obtain from traditional models or statistical analysis (Merz et al., 2021). Based on the powerful nonlinear fitting ability of deep learning for time series, we use interpretable methods to break the black box characteristics. In this way, hydrological knowledge is obtained from models trained on large samples, revealing the spatiotemporal characteristics of the

10.1029/2024EF004751

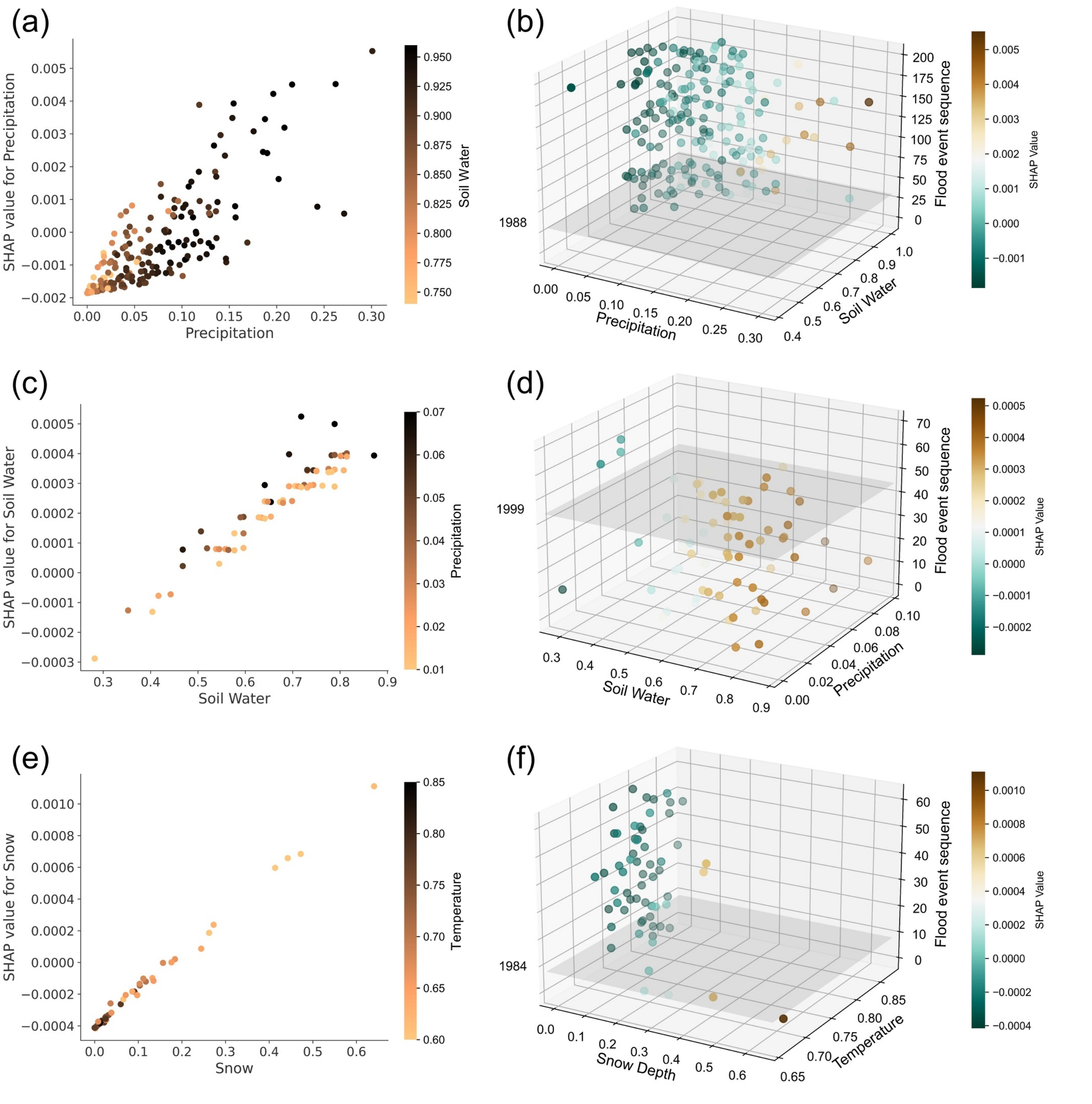

Figure 10. (a, c, and e) SHAP values for each flood event of pairwise variables with the most significant interactions in three representative watersheds, reflecting their impact on floods at different values of the variables. (b, d, and f) The dynamic impact of pairwise variables on floods over time, colored by SHAP values. The *x* and *y* axes represent the range of variable values, and the *z*-axis represents the chronological order of flood events. The auxiliary planes shown in gray indicate the timing of abrupt changes in the dominant driving factors within the watersheds.

dominant driving factors of floods, the interactive responses to these factors in flood formation and the mutations of different flood generation mechanisms.

More realistic flood simulations appear to be beneficial for reverse derivation to increase hydrological insights, especially for accurate restoration of flood peaks (Jiang, Bevacqua, & Zscheischler, 2022). Informer provides more reliable estimates of peak values in time regions than other deep learning models. The addition of the sparse attention structure allows the Informer to focus on the points that really contribute to the target value in the long-term sequence. This is due to the fact that the Informer calculates the activity and KL divergence of each query compared to the traditional Transformer, thus reducing the number of low-contribution pairs (L. Shen & Wang, 2022; Zhou et al., 2023). The special structure enables Informer to be more sensitive to drastic changes in flow, even if longer time steps are considered in the simulation, which has been confirmed in some recent water temperature and power system simulation studies (Gong et al., 2022; Huang & Jiang, 2022; Wang et al., 2022). Another issue that needs to be considered is multicollinearity of multiple model input factors. Jiang, Bevacqua, and Zscheischler (2022) pointed out that it is difficult to assign high importance to highly correlated variables at the same time because the interpretation technique serves the model rather than fully conforming to the real world. An Informer that can accurately identify the contribution of variables to floods may make the description of dynamic driving factors in the explanation more realistic and improve the difficulties caused by multicollinearity.

Interpretable technology explores the internal working principles of deep learning and derives the dominant driving factors of floods in large sample watersheds. The results we reveal are consistent with past research (Berghuijs et al., 2019; Brunner et al., 2020; Stein et al., 2020; M. Shen & Chui, 2023). Berghuijs et al. (2016) used a hypothesis testing process to obtain the main drivers of floods in the United States, which is based on simple formulas of meteorological and hydrological variables corresponding to different flood generation mechanisms. The spatial distribution of precipitation-dominated and snowmelt-dominated watersheds was jointly confirmed (Berghuijs et al., 2016). The difference in this study is that we used the data side to discover the physical laws contained in big data, and we believe that the latter can adapt to changing climate and underlying surface conditions. The significance for scientists is that they only need to understand basic hydrological knowledge (such as how melting snow flows into rivers to cause flow changes), and they can obtain reasonable explanations for complex phenomena between regions through this process of reverse derivation. Through some inferences, an understanding is obtained that is close to or completely consistent with the fieldwork. It is worth noting that we take sufficient time steps into account when calculating the driving factors of floods, because floods may not only be controlled by sharp precipitation, but also by abundant soil water over long periods of time in the past. It turns out that factors in the 6 days before a flood play different levels of importance, depending on the timing of their involvement in hydrological processes (Bloeschl et al., 2017; Xiang et al., 2020). In watersheds dominated by different primary driving factors, the contributions of factors within the time steps also vary. Precipitation-dominated watersheds are primarily influenced by factors from the previous day, whereas snow-dominated watersheds are affected more significantly by factors from the past 2–6 days Shen et al. (2023) used a data-driven tree simplified model to discuss the flood generation mechanism. In contrast, we use deep learning to extend the variables in time and restore the realistic flood process as much as possible. This has also become a more accurate, comprehensive and even unprecedented method to reveal the flood process. Efficient access to undiscovered hydrological information.

We computed the results of the interpretability technique through a sliding window to reveal changes in the dominant driving factors. Although only a minority of watersheds exhibited abrupt changes in dominant factors, these changes are still significant over a long period. These abrupt changes in the United States show significant regional differences in timing and trends (Milly et al., 2008; Min et al., 2011; Trenberth et al., 2003). Floods caused by precipitation have become more predominant and vary greatly in magnitude, which may be related to atmospheric warming, that is, increased atmospheric humidity leading to more extreme precipitation events (Arnell & Gosling, 2016; Pall et al., 2011). The likelihood of extreme rainfall under climate warming increases, meaning that every decade, more floods in watersheds are caused by precipitation, with these changes concentrated in the 1990s. Conversely, the frequency of floods caused by soil water is decreasing at a slow rate, with the phenomenon of soil water surplus triggered by prior precipitation decreasing, which is also related to the increased frequency of rapid and sharp precipitation events. Jiang, Bevacqua, and Zscheischler (2022) found similar patterns in European regions, suggesting that the changing trends in flood mechanisms revealed by this study may be cross-regional or even global.

The formation of floods is often influenced by a combination of factors, and the interaction between variables over time and how they affect river floods need to be revealed (Merz et al., 2021; Slater et al., 2021). Floods may become more severe if multiple processes and conditions favorable to high river flows occur simultaneously (Bloeschl et al., 2017; Sikorska et al., 2015). Through interpretable deep learning, we have revealed the time-varying interactions between meteorological and hydrological variables in watersheds, with the distribution of variable values visualized within these interactions. Prolonged intense precipitation and prior high soil moisture often act together, with excessive net radiation and temperatures acting on snowmelt to cause flood events (Bertola et al., 2021; Sharma et al., 2018; Westra et al., 2013). The driving factors of floods are not extreme in themselves; their combined effects can also trigger extreme or unprecedented flood events, posing significant challenges to flood risk management (Winsemius et al., 2016; Zscheischler et al., 2018). Phenomena that have not yet been defined by physical formulas exist in complex natural processes, especially in the hydrological cycle. The interpretable deep learning approach proposed in this study provides a means of data mining to reveal the complex driving factors of floods, which is crucial for improving current flood risk strategies and developing new ones for the future.

A better understanding of the dynamic driving factors of river floods is crucial for explaining past flood variations and improving future flood risk predictions (AghaKouchak et al., 2020; Hall et al., 2014; Sterling et al., 2013). Deep learning models, with their capacity to capture complex nonlinear relationships between variables, have become an effective tool for hydrological simulation. However, what's more valuable is the ability to mine insights from robust feature fittings to enhance our understanding of hydrological knowledge (C. Shen et al., 2023; Tsai et al., 2021). Employing interpretability techniques on large-sample hydrometeorological variables with temporal attributes and causal relationships allows for the consideration of the entire process of flood formation. Xu et al. (2022) demonstrated how deep learning models for time series prediction make decisions about future flows, with sliding windows of a time step length being crucial for each moment's prediction. This is reflected in the SHAP calculations of this study, which take into account past information for future flow predictions, ensuring that the mechanism of any variable that had an impact in the past can be quantified at each moment. Interpretable methods provide a pathway to a deeper understanding of how machine learning works, disentangling relationships between attributes or variables and enabling people to comprehend them. This is a data-driven reverse deduction approach that generates new insights into hydrological features. It has the potential to be extended to more fields in the future to reveal underlying natural physical processes, such as revealing how human activities under carbon emission scenarios affect regional hydrological cycles, identifying which factors truly influence the increase in precipitation in urbanized areas, and understanding how vegetation responds to moisture forcing. The proposed technical framework offers a feasible approach to deriving underlying physical laws from big data and enriching our knowledge base.

5. Conclusions

The formation of floods is often triggered under the complex influence of multiple variables, occurring in parallel across different spatial and temporal scales. To reveal the dynamic driving factors of floods, we propose a new interpretable framework based on time series deep learning that quantifies the response of the flood generation process to hydrometeorological variables at different times. Precise flood simulation is a prerequisite for the application of interpretability techniques, and the Informer model, with its sparse attention mechanism sensitive to extreme values, was chosen for pre-training on a large sample. Under the same 10-fold cross-validation, the Informer achieved more robust and reliable flood peak simulation results compared to other deep learning models (with NSE exceeding 0.6 in 70% of the watersheds). We used SHAP to explain the decision-making process of the Informer, identifying three main patterns of flood triggers across 482 US watersheds, corresponding to dominant driving factors: precipitation, excess soil water, and snowmelt. Precipitation is primarily found along the western coast and in eastern inland areas, while excess soil water is more prevalent in central regions. The spatial distribution of dominant driving factors reflects changes in watershed geography and climate characteristics, consistent with findings from earlier studies. High values of precipitation, soil water, and snow depth all contribute to flood generation. Notably, variables from different times before a flood can have significantly different impacts, reflected in opposing mechanisms (either promoting or inhibiting), and it is not always the variables closest to the flood event that play a crucial role. In precipitation-dominated watersheds, precipitation significantly controls flood occurrences, particularly in the western coastal and eastern inland regions, with the primary influence stemming from precipitation within the past day. In contrast, in soil water and snow-dominated watersheds, the dominant factors are more pronounced in the southern coastal and Rocky Mountain regions. The generation of floods is not entirely influenced by the conditions of the past day, but more by the conditions of the past 2–6 days. We further discovered that over 40% of the watersheds experienced abrupt changes in the dominant driving factors for floods during 1981–2020, with watersheds dominated by precipitation undergoing more significant changes, consistent with responses to climate change. Furthermore, different values of variables lead to varying interactive effects, which also manifest as diverse dynamic effects on flood events over time. Advanced interpretable deep learning, through this reverse deduction approach, by exploring the inner workings of pre-trained deep learning models, uncovers potential hydrological information over long sequences. Importantly, it transforms deep learning from merely being a black box tool for fitting nonlinear relationships into an effective means for gaining new insights into hydrological features, helping hydrologists better understand natural physical processes.

Data Availability Statement

The Carvan data set is derived from Kratzert et al. (2024), which includes daily hydro-meteorological time series data for 482 watersheds across the United States. The final data retrieval date was February 2024. The deep learning Informer code used for framework modeling in the study is sourced from Zhou et al. (2021) and available at GitHub (https://github.com/zhouhaoyi/Informer2020). The code for the interpretability technique SHAP is sourced from S. Lundberg and Lee (2017) and available at GitHub (https://github.com/shap/shap).

Acknowledgments

The authors would like to acknowledge the financial support received from Projects of National Key R&D Program of China (2022YFC3002903), Guangdong Basic and Applied Basic Research Foundation (2023B1515040028), Guangdong Provincial Bureau of Hydrology (440001-2023-10716), National Natural Science Foundation of China (U2243219), and the Research Council of Norway (FRINATEK Project 274310, KLIMAFORSK Project 302457).

References

Aas, K., Jullum, M., & Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artificial Intelligence*, *298*, 103502. https://doi.org/10.1016/j.artint.2021.103502

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, *54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., et al. (2020). Climate extremes and compound hazards in a warming world. In R. Jeanloz & K. H. Freeman (Eds.), *Annual review of earth and planetary sciences* (Vol. 48(1), pp. 519–548). https://doi.org/10.1146/annurev-earth-071719-055228

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., et al. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, *5*(2), 171–182. https://doi.org/10.1002/2016EF000485

Anderson, B. J., Slater, L. J., Dadson, S. J., Blum, A. G., & Prosdocimi, I. (2022). Statistical attribution of the influence of urban and tree cover change on streamflow: A comparison of large sample statistical approaches. *Water Resources Research*, *58*(5), e2021WR030742. https://doi.org/10.1029/2021WR030742

Anderson, S., & Radic, V. (2022). Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. *Hydrology and Earth System Sciences*, *26*(3), 795–825. https://doi.org/10.5194/hess-26-795-2022

Archfield, S. A., Hirsch, R. M., Viglione, A., & Bloeschl, G. (2016). Fragmented patterns of flood change across the United States. *Geophysical Research Letters*, *43*(19), 10232–10239. https://doi.org/10.1002/2016GL070590

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climate Change*, *134*(3), 387–401. https://doi.org/10.1007/s10584-014-1084-5

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, *57*(2), e2020WR028673. https://doi.org/10.1029/2020WR028673

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The relative importance of different flood-generating mechanisms across Europe. *Water Resources Research*, *55*(6), 4582–4593. https://doi.org/10.1029/2019WR024841

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, *43*(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Bertola, M., Viglione, A., Vorogushyn, S., Lun, D., Merz, B., & Bloeschl, G. (2021). Do small and large floods have the same drivers of change? A regional attribution analysis in Europe. *Hydrology and Earth System Sciences*, *25*(3), 1347–1364. https://doi.org/10.5194/hess-25-1347-2021

Bian, L., Qin, X., Zhang, C., Guo, P., & Wu, H. (2023). Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM-a case study for runoff simulation in an arid area. *Journal of Hydrology*, *625*, 130091. https://doi.org/10.1016/j.jhydrol.2023.130091

Bloeschl, G., Hall, J., Parajka, J., Perdigao, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Bloeschl, G., Hall, J., Viglione, A., Perdigao, R. A. P., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*(7772), 108–111. https://doi.org/10.1038/s41586-019-1495-6

Brunner, M., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future streamflow regime changes in the United States: Assessment using functional classification. *Hydrology and Earth System Sciences*, *24*(8), 3951–3966. https://doi.org/10.5194/hess-24-3951-2020

Cheng, S., Cheng, L., Qin, S., Zhang, L., Liu, P., Liu, L., et al. (2022). Improved understanding of how catchment properties control hydrological partitioning through machine learning. *Water Resources Research*, *58*(4), e2021WR031412. https://doi.org/10.1029/2021WR031412

Davenport, F. V., Herrera-Estrada, J. E., Burke, M., & Diffenbaugh, N. S. (2020). Flood size increases nonlinearly across the western United States in response to lower snow-precipitation ratios. *Water Resources Research*, *56*(1), e2019WR025571. https://doi.org/10.1029/2019WR025571

De la Fuente, L. A., Ehsani, M. R., Gupta, H. V., & Condon, L. E. (2024). Toward interpretable LSTM-based modeling of hydrological systems. *Hydrology and Earth System Sciences*, *28*(4), 945–971. https://doi.org/10.5194/hess-28-945-2024

Dethier, E. N., Sartain, S. L., Renshaw, C. E., & Magilligan, F. J. (2020). Spatially coherent regional changes in seasonal extreme streamflow events in the United States and Canada since 1950. *Science Advances*, *6*(49), eaba5939. https://doi.org/10.1126/sciadv.aba5939

Do, H. X., Mei, Y., & Gronewold, A. D. (2020). To what extent are changes in flood magnitude related to changes in precipitation extremes? *Geophysical Research Letters*, *47*(18), e2020GL088684. https://doi.org/10.1029/2020GL088684

Do, H. X., Westra, S., & Leonard, M. (2017). A global-scale investigation of trends in annual maximum streamflow. *Journal of Hydrology*, *552*, 28–43. https://doi.org/10.1016/j.jhydrol.2017.06.015

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, *6*(5), 508–513. https://doi.org/10.1038/NCLIMATE2941

Feng, D., Beck, H., Lawson, K., & Shen, C. (2023). The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. *Hydrology and Earth System Sciences*, *27*(12), 2357–2373. https://doi.org/10.5194/hess-27-2357-2023

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, *56*(9), e2019WR026793. https://doi.org/10.1029/2019WR026793

Fong, R. C., & Vedaldi, A. (2017). *Interpretable explanations of black boxes by meaningful perturbation* (pp. 3449–3457). International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2017.371

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., et al. (2022). Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences*, *26*(13), 3377–3392. https://doi.org/10.5194/hess-26-3377-2022

Gong, M., Zhao, Y., Sun, J., Han, C., Sun, G., & Yan, B. (2022). Load forecasting of district heating system based on Informer. *Energy*, *253*, 124179. https://doi.org/10.1016/j.energy.2022.124179

Guo, D., Westra, S., & Maier, H. R. (2017). Use of a scenario-neutral approach to identify the key hydro-meteorological attributes that impact runoff from a natural catchment. *Journal of Hydrology*, *554*, 317–330. https://doi.org/10.1016/j.jhydrol.2017.09.021

Hall, J., Arheimer, B., Borga, M., Brazdil, R., Claps, P., Kiss, A., et al. (2014). Understanding flood regime changes in Europe: A state-of-the-art assessment. *Hydrology and Earth System Sciences*, *18*(7), 2735–2772. https://doi.org/10.5194/hess-18-2735-2014

Han, J., Liu, Z., Woods, R., McVicar, T. R., Yang, D., Wang, T., et al. (2024). Streamflow seasonality in a snow-dwindling world. *Nature*, *629*(8014), 1075–1081. https://doi.org/10.1038/s41586-024-07299-y

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., et al. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821. https://doi.org/10.1038/NCLIMATE1911

Huang, X., & Jiang, A. (2022). Wind power generation forecast based on multi-step informer network. *Energies*, *15*(18), 6642. https://doi.org/10.3390/en15186642

Jiang, S., Bevacqua, E., & Zscheischler, J. (2022). River flooding mechanisms and their changes in Europe revealed by explainable machine learning. *Hydrology and Earth System Sciences*, *26*(24), 6339–6359. https://doi.org/10.5194/hess-26-6339-2022

Jiang, S., Tarasova, L., Yu, G., & Zscheischler, J. (2024). Compounding effects in flood drivers challenge estimates of extreme river floods. *Science Advances*, *10*(13), eadl4005. https://doi.org/10.1126/sciadv.adl4005

Jiang, S., Zheng, Y., Wang, C., & Babovic, V. (2022). Uncovering flooding mechanisms across the contiguous United States through interpretive deep learning on representative catchments. *Water Resources Research*, *58*(1), e2021WR030185. https://doi.org/10.1029/2021WR030185

Jing, H., He, X., Tian, Y., Lancia, M., Cao, G., Crivellari, A., et al. (2023). Comparison and interpretation of data-driven models for simulating site-specific human-impacted groundwater dynamics in the North China Plain. *Journal of Hydrology*, *616*, 128751. https://doi.org/10.1016/j.jhydrol.2022.128751

Keller, L., Rossler, O., Martius, O., & Weingartner, R. (2018). Delineation of flood generating processes and their hydrological response. *Hydrological Processes*, *32*(2), 228–240. https://doi.org/10.1002/hyp.11407

Kiptum, A., Antonarakis, A. S., Todd, M. C., & Guigma, K. (2024). Characteristics, drivers, and predictability of flood events in the Tana River Basin, Kenya. *Journal of Hydrology: Regional Studies*, *53*, 101748. https://doi.org/10.1016/j.ejrh.2024.101748

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. *Water Resources Research*, *56*(9), e2019WR025975. https://doi.org/10.1029/2019WR025975

Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Koya, S. R., & Roy, T. (2023). Temporal fusion transformers for streamflow prediction: Value of combining attention with recurrence. *arXiv.Org*. https://doi.org/10.48550/arXiv.2305.12335

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2021). A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences*, *25*(5), 2685–2703. https://doi.org/10.5194/hess-25-2685-2021

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., et al. (2024). Caravan - a global community dataset for large-sample hydrology (1.4) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.10968468

Kroner, A., Senden, M., Driessens, K., & Goebel, R. (2020). Contextual encoder-decoder network for visual saliency prediction. *Neural Networks*, *129*, 261–270. https://doi.org/10.1016/j.neunet.2020.05.004

Lang, M., Ouarda, T., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*(3–4), 103–117. https://doi.org/10.1016/S0022-1694(99)00167-5

Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., et al. (2022). Hydrological concept formation inside long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *26*(12), 3079–3101. https://doi.org/10.5194/hess-26-3079-2022

Li, L., Qiao, J., Yu, G., Wang, L., Li, H.-Y., Liao, C., & Zhu, Z. (2022). Interpretable tree-based ensemble model for predicting beach water quality. *Water Research*, *211*, 118078. https://doi.org/10.1016/j.watres.2022.118078

Liu, J., Bian, Y., & Shen, C. (2023). Probing the limit of hydrologic predictability with the Transformer network. *arXiv.Org*. https://doi.org/10.48550/arXiv.2306.12384

Liu, J., Feng, S., Gu, X., Zhang, Y., Beck, H. E., Zhang, J., & Yan, S. (2022). Global changes in floods and their drivers. *Journal of Hydrology*, *614*, 128553. https://doi.org/10.1016/j.jhydrol.2022.128553

Liu, Q., Gui, D., Zhang, L., Niu, J., Dai, H., Wei, G., & Hu, B. X. (2022). Simulation of regional groundwater levels in arid regions using interpretable machine learning models. *Science of the Total Environment*, *831*, 154902. https://doi.org/10.1016/j.scitotenv.2022.154902

Liu, Y., Duffy, K., Dy, J. G., & Ganguly, A. R. (2023). Explainable deep learning for insights in El Nino and river flows. *Nature Communications*, *14*(1), 339. https://doi.org/10.1038/s41467-023-35968-5

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions [Software]. *arXiv*. https://doi.org/10.48550/arXiv.1705.07874

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lusch, B., Kutz, J. N., & Brunton, S. L. (2018). Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, *9*(1), 4950. https://doi.org/10.1038/s41467-018-07210-0

Machado, M. J., Botero, B. A., Lopez, J., Frances, F., Diez-Herrero, A., & Benito, G. (2015). Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrology and Earth System Sciences*, *19*(6), 2561–2576. https://doi.org/10.5194/hess-19-2561-2015

Mallakpour, I., & Villarini, G. (2015). The changing nature of flooding across the central United States. *Nature Climate Change*, *5*(3), 250–254. https://doi.org/10.1038/NCLIMATE2516

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2023). Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems*, *2*(1). https://doi.org/10.1175/AIES-D-22-0058.1

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., et al. (2022). Understanding heavy tails of flood peak distributions. *Water Resources Research*, *58*(6), e2021WR030506. https://doi.org/10.1029/2021WR030506

Merz, B., Bloesch, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., et al. (2021). Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth and Environment*, *2*(9), 592–609. https://doi.org/10.1038/s43017-021-00195-3

Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., & Hundecha, Y. (2012). HESS Opinions More efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrology and Earth System Sciences*, *16*(5), 1379–1387. https://doi.org/10.5194/hess-16-1379-2012

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change - stationarity is dead: Whither water management? *Science*, *319*(5863), 573–574. https://doi.org/10.1126/science.1151915

Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378–381. https://doi.org/10.1038/nature09763

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences* (Vol. 116(44), 22071–22080). https://doi.org/10.1073/pnas.1900654116

Musselman, K. N., Lehner, F., Ikeda, K., Clark, M. P., Prein, A. F., Liu, C., et al. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. *Nature Climate Change*, *8*(9), 808–812. https://doi.org/10.1038/s41558-018-0236-4

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

Ozaki, Y., Tanigaki, Y., Watanabe, S., & Onishi, M. (2020). Multiobjective tree-structured parzen estimator for computationally expensive optimization problems. In *Gecco '20: Proceedings of the 2020 genetic and evolutionary computation conference* (pp. 533–541). https://doi.org/10.1145/3377930.3389817

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G. J., et al. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, *470*(7334), 382–385. https://doi.org/10.1038/nature09762

Regonda, S. K., Rajagopalan, B., Clark, M., & Pitlick, J. (2005). Seasonal cycle shifts in hydroclimatology over the western United States. *Journal of Climate*, *18*(2), 372–384. https://doi.org/10.1175/JCLI-3272.1

Rentschler, J., Avner, P., Marconcini, M., Su, R., Strano, E., Vousdoukas, M., & Hallegatte, S. (2023). Global evidence of rapid urban growth in flood zones since 1985. *Nature*, *622*(7981), 87–92. https://doi.org/10.1038/s41586-023-06468-9

Rong, Y., Bates, P., Neal, J., Archer, L., Hatchard, S., & Kendon, E. (2024). Impact of soil moisture dynamics and precipitation pattern on UK urban pluvial flood hazards under climate change. *Earth's Future*, *12*(5), e2023EF004073. https://doi.org/10.1029/2023EF004073

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If precipitation extremes are increasing, why aren't floods? *Water Resources Research*, *54*(11), 8545–8551. https://doi.org/10.1029/2018WR023749

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth and Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Shen, L., & Wang, Y. (2022). Tcct: Tightly-coupled convolutional transformer on time series forecasting. *Neurocomputing*, *480*, 131–145. https://doi.org/10.1016/j.neucom.2022.01.039

Shen, M., & Chui, T. F. M. (2023). Quantifying the relative contributions of different flood generating mechanisms to floods across CONUS. *Journal of Hydrology*, *626*, 130255. https://doi.org/10.1016/j.jhydrol.2023.130255

Sikorska, A. E., Viviroli, D., & Seibert, J. (2015). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resources Research*, *51*(10), 7959–7976. https://doi.org/10.1002/2015WR017326

Singh, N. K., & Basu, N. B. (2022). The human factor in seasonal streamflows across natural and managed watersheds of North America. *Nature Sustainability*, *5*(5), 397–405. https://doi.org/10.1038/s41893-022-00848-1

Slater, L. J., Anderson, B., Buechel, M., Dadson, S., Han, S., Harrigan, S., et al. (2021). Nonstationary weather and water extremes: A review of methods for their detection, attribution, and management. *Hydrology and Earth System Sciences*, *25*(7), 3897–3935. https://doi.org/10.5194/hess-25-3897-2021

Slater, L. J., & Villarini, G. (2016). Recent trends in US flood risk. *Geophysical Research Letters*, *43*(24), 12428–12436. https://doi.org/10.1002/2016GL071199

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F., & Woods, R. A. (2021). How Do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. *Water Resources Research*, *57*(4), e2020WR028300. https://doi.org/10.1029/2020WR028300

Stein, L., Pianosi, F., & Woods, R. (2020). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, *34*(7), 1514–1529. https://doi.org/10.1002/hyp.13678

Sterling, S. M., Ducharne, A., & Polcher, J. (2013). The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, *3*(4), 385–390. https://doi.org/10.1038/NCLIMATE1690

Stojic, A., Stanic, N., Vukovic, G., Stanisic, S., Perisic, M., Sostaric, A., & Lazic, L. (2019). Explainable extreme gradient boosting tree-based prediction of toluene, ethylbenzene and xylene wet deposition. *Science of the Total Environment*, *653*, 140–147. https://doi.org/10.1016/j.scitotenv.2018.10.368

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665. https://doi.org/10.1007/s10115-013-0679-x

Tarasova, L., Basso, S., Wendi, D., Viglione, A., Kumar, R., & Merz, R. (2020). A process-based framework to characterize and classify runoff events: The event typology of Germany. *Water Resources Research*, *56*(5), e2019WR026951. https://doi.org/10.1029/2019WR026951

Tarasova, L., Lun, D., Merz, R., Blöschl, G., Basso, S., Bertola, M., et al. (2023). Shifts in flood generation processes exacerbate regional flood anomalies in Europe. *Communications Earth & Environment*, *4*(1), 1–12. https://doi.org/10.1038/s43247-023-00714-8

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w

Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, *84*(9), 1205–1217. https://doi.org/10.1175/BAMS-84-9-1205

Tripathy, K. P., & Mishra, A. K. (2024). Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. *Journal of Hydrology*, *628*, 130458. https://doi.org/10.1016/j.jhydrol.2023.130458

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

Turkington, T., Breinl, K., Ettema, J., Alkema, D., & Jetten, V. (2016). A new flood type classification method for use in climate change impact studies. *Weather and Climate Extremes*, *14*, 1–16. https://doi.org/10.1016/j.wace.2016.10.001

Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., & Wong, W. K. (2016). Evidence for changes in the magnitude and frequency of observed rainfall vs. Snowmelt driven floods in Norway. *Journal of Hydrology*, *538*, 33–48. https://doi.org/10.1016/j.jhydrol.2016.03.066

Wang, H., Liu, J., Klaar, M., Chen, A., Gudmundsson, L., & Holden, J. (2024). Anthropogenic climate change has influenced global river flow seasonality. *Science*, *383*(6686), 1009–1014. https://doi.org/10.1126/science.adi9501

Wang, H.-K., Song, K., & Cheng, Y. (2022). A hybrid forecasting model based on CNN and informer for short-term wind power. *Frontiers in Energy Research*, *9*, 788320. https://doi.org/10.3389/fenrg.2021.788320

Wasko, C., & Nathan, R. (2019). Influence of changes in rainfall and soil moisture on trends in flooding. *Journal of Hydrology*, *575*, 432–441. https://doi.org/10.1016/j.jhydrol.2019.05.054

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *Journal of Climate*, *26*(11), 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Whitfield, P. H., Burn, D. H., Hannaford, J., Higgins, H., Hodgkins, G. A., Marsh, T., & Looser, U. (2012). Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. *Hydrological Sciences Journal*, *57*(8), 1562–1579. https://doi.org/10.1080/02626667.2012.728706

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., et al. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381–385. https://doi.org/10.1038/nclimate2893

Xiang, Z., Yan, J., & Demir, I. (2020). A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resources Research*, *56*(1), e2019WR025326. https://doi.org/10.1029/2019WR025326

Xu, Y., Hu, C., Wu, Q., Jian, S., Li, Z., Chen, Y., et al. (2022). Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *Journal of Hydrology*, *608*, 127553. https://doi.org/10.1016/j.jhydrol.2022.127553

Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., & Ran, G. (2023). Deep transfer learning based on transformer for flood forecasting in data-sparse watersheds. *Journal of Hydrology*, *625*, 129956. https://doi.org/10.1016/j.jhydrol.2023.129956

Yin, J., Guo, S., Gentine, P., Sullivan, S. C., Gu, L., He, S., et al. (2021). Does the hook structure constrain future flood intensification under anthropogenic climate warming? *Water Resources Research*, *57*(2), e2020WR028491. https://doi.org/10.1029/2020WR028491

Zhang, C., Wang, L., Cheng, S., & Li, Y. (2022). SwinSUNet: Pure transformer network for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 5224713. https://doi.org/10.1109/TGRS.2022.3160007

Zhang, J., Zhu, Y., Zhang, X., Ye, M., & Yang, J. (2018). Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*, *561*, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065

Zhang, S., Zhou, L., Zhang, L., Yang, Y., Wei, Z., Zhou, S., et al. (2022). Reconciling disagreement on global river flood changes in a warming climate. *Nature Climate Change*, *12*(12), 1160–1167. https://doi.org/10.1038/s41558-022-01539-7

Zhang, X. S., Amirthanathan, G. E., Bari, M. A., Laugesen, R. M., Shin, D., Kent, D. M., et al. (2016). How streamflow has changed across Australia since the 1950s: Evidence from the network of hydrologic reference stations. *Hydrology and Earth System Sciences*, *20*(9), 3947–3965. https://doi.org/10.5194/hess-20-3947-2016

Zhou, H., Li, J., Zhang, S., Zhang, S., Yan, M., & Xiong, H. (2023). Expanding the prediction capacity in long sequence time-series forecasting. *Artificial Intelligence*, *318*, 103886. https://doi.org/10.1016/j.artint.2023.103886

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting [Software]. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3